

Analisis Kinerja Model Random Forest dengan Teknik Manhattan-SMOTE pada Deteksi Fraud Transaksi Kartu Kredit Imbalance

Bq Nadila Nuzululnisa, Hairani Hairani

Universitas Bumigora, Mataram, Indonesia

Correspondence : e-mail: bnadilanzululnisa@gmail.com

Abstrak

Ketidakeimbangan data merupakan salah satu tantangan utama dalam pengembangan sistem deteksi penipuan transaksi kartu kredit. Model pembelajaran mesin cenderung bias terhadap kelas mayoritas, sehingga sulit mendeteksi transaksi fraud yang tergolong sebagai kelas minoritas. Penelitian ini bertujuan untuk meningkatkan kinerja deteksi fraud dengan mengimplementasikan teknik oversampling Manhattan-SMOTE sebagai solusi penyeimbang data sebelum pelatihan model Random Forest. Manhattan-SMOTE merupakan pengembangan dari metode SMOTE konvensional yang menggunakan jarak Manhattan dalam proses interpolasi data sintesis, sehingga lebih stabil dan akurat untuk data berdimensi tinggi. Hasil evaluasi menunjukkan bahwa model Random Forest tanpa oversampling menghasilkan akurasi 81.18% dengan recall yang rendah, yaitu 36.26%. Setelah diterapkan Manhattan-SMOTE, nilai recall meningkat menjadi 67%, F1-score menjadi 0.50, dan ROC AUC melonjak dari 0.75 menjadi 0.96, meskipun akurasi menurun menjadi 70%. Hasil ini menunjukkan bahwa teknik Manhattan-SMOTE secara signifikan meningkatkan kemampuan model dalam mengenali transaksi penipuan tanpa mengorbankan performa klasifikasi secara keseluruhan. Kombinasi Random Forest dan Manhattan-SMOTE terbukti efektif dalam mengatasi ketidakseimbangan kelas dan layak diterapkan dalam sistem deteksi fraud berbasis kecerdasan buatan.

Kata kunci: deteksi fraud, data tidak seimbang, random forest, manhattan-SMOTE, transaksi kartu kredit.

Abstract

Data imbalance is one of the main challenges in developing credit card fraud detection systems. Machine learning models tend to be biased toward the majority class, making it difficult to detect fraud transactions, which belong to the minority class. This study aims to improve fraud detection performance by implementing the Manhattan-SMOTE oversampling technique as a data balancing solution prior to training the Random Forest model. Manhattan-SMOTE is an enhancement of the conventional SMOTE method that uses Manhattan distance during the interpolation of synthetic samples, resulting in more stable and accurate outcomes for high-dimensional data. Evaluation results show that the Random Forest model without oversampling achieved an accuracy of 81.18% but had a low recall of only 36.26%. After applying Manhattan-SMOTE, the recall increased to 67%, the F1-score reached 0.50, and the ROC AUC rose from 0.75 to 0.96, although the accuracy declined to 70%. These results indicate that the Manhattan-SMOTE technique significantly improves the model's ability to detect fraudulent transactions without compromising overall classification performance. The combination of Random Forest and Manhattan-SMOTE is proven to be effective in addressing class imbalance and is suitable for implementation in artificial intelligence-based fraud detection systems.

Keywords: fraud detection, imbalanced data, random forest, manhattan-SMOTE, credit card transaction.

1. Pendahuluan

Dalam era kemajuan teknologi digital dan sistem pembayaran elektronik, ada peningkatan yang signifikan dalam penggunaan kartu kredit sebagai sarana transaksi. [1] Namun, dengan pertumbuhan ini juga muncul kejahatan siber yang semakin meningkat, terutama penipuan transaksi yang berkaitan dengan kartu kredit. [2] Tindakan penipuan ini tidak hanya merugikan individu tetapi juga lembaga keuangan

dengan kerugian yang besar, sehingga kemampuan untuk mendeteksi penipuan secara otomatis dan cepat menjadi hal yang sangat penting dalam dunia keuangan. [3] Salah satu tantangan utama dalam mendeteksi penipuan jenis ini adalah data yang sangat tidak seimbang, di mana jumlah transaksi penipuan jauh lebih sedikit dibandingkan dengan transaksi yang sah. [4] Kondisi ini menyebabkan algoritma klasifikasi sering kali mengabaikan kelas yang lebih kecil, yang justru merupakan sasaran penting untuk deteksi. [5]

Berbagai metode telah dikembangkan untuk menangani ketidakseimbangan ini, salah satunya adalah *Synthetic Minority Over-sampling Technique* (SMOTE), yang berfungsi dengan cara menciptakan sampel sintetis dari kelas minoritas agar distribusi data menjadi lebih seimbang. [6] Metode SMOTE yang tradisional menggunakan *Euclidean Distance* untuk menghasilkan sampel sintetis, tetapi pendekatan ini kurang efektif ketika berhadapan dengan data yang memiliki banyak dimensi atau distribusi yang non-linear. [7] Sebagai opsi lain, *Manhattan Distance* dipilih karena kemampuannya yang lebih stabil terhadap data yang memiliki *outlier* serta lebih cocok untuk data dengan distribusi tidak merata. [8] Oleh sebab itu, penelitian ini menggunakan Manhattan-SMOTE, yaitu varian dari SMOTE yang mengganti *Euclidean Distance* dengan *Manhattan Distance*, untuk meningkatkan kualitas pembuatan data sintetis dan kinerja model klasifikasi dalam mendeteksi penipuan. [9]

Dalam penelitian ini, model klasifikasi yang diterapkan adalah Random Forest, yang dikenal efektif dalam menangani data kompleks, mampu mengatasi *outlier*, dan tetap stabil dari risiko *overfitting*. [10] Tujuan dari penelitian ini adalah menguji bagaimana teknik Manhattan-SMOTE dapat memperbaiki kinerja model Random Forest dalam mendeteksi transaksi kartu kredit yang curang. Metrik yang digunakan untuk evaluasi mencakup akurasi, precision, recall, F1-score, AUC-ROC, dan PR-AUC, yang memberikan gambaran yang lebih lengkap mengenai kinerja model.

Manfaat dari penelitian ini meliputi penyediaan metode deteksi penipuan yang lebih akurat, membantu lembaga keuangan untuk mengurangi kerugian akibat penipuan, dan memberikan kontribusi ilmiah dalam pengembangan teknik penyeimbangan data untuk pembelajaran mesin. Hasil penelitian ini juga dapat menjadi referensi bagi pengambilan keputusan dalam merancang sistem keamanan untuk transaksi keuangan berbasis machine learning.

Salah satu gap dalam penelitian yang ingin dijawab oleh studi ini adalah minimnya penerapan *Manhattan Distance* dalam proses oversampling untuk deteksi penipuan dan kurangnya penelitian yang membandingkan kinerja Manhattan-SMOTE dengan SMOTE tradisional dalam konteks model Random Forest. [11] Sebagian besar penelitian sebelumnya hanya berfokus pada perbandingan antara model klasifikasi atau hanya menerapkan SMOTE standard tanpa mempertimbangkan jenis jarak yang digunakan dalam proses *oversampling*. Dengan memanfaatkan pendekatan Manhattan-SMOTE, penelitian ini menghadirkan kebaruan melalui teknik *resampling* yang lebih sesuai dan akurat untuk data yang sangat tidak seimbang, serta menghasilkan model yang lebih efisien dalam mendeteksi penipuan kartu kredit. [12]

Ada berbagai penelitian sebelumnya yang telah mencoba mendeteksi penipuan dengan pendekatan yang serupa. Contohnya, penelitian oleh Carcillo et al. yang mengaplikasikan pembelajaran *ensemble* dan SMOTE untuk menemukan penipuan dalam sektor perbankan, serta Chawla et al. yang pertama kali memperkenalkan SMOTE untuk menyeimbangkan data dalam klasifikasi. Walau begitu, masih sedikit penelitian yang secara khusus menggunakan varian Manhattan-SMOTE dalam Random Forest. [13] Penyelidikan oleh Liu et al. menunjukkan bahwa Jarak Manhattan lebih unggul dalam menangani kebisingan saat membuat data sintetis. Oleh karena itu, pendekatan yang diambil dalam studi ini diharapkan dapat menjadi solusi untuk tantangan yang dihadapi dalam mengklasifikasikan data penipuan yang sangat tidak seimbang. [14] Selain itu, Random Forest merupakan algoritma *ensemble* berbasis pohon keputusan yang telah terbukti efektif dalam menangani data dengan fitur non-linear dan ketidakseimbangan kelas, dan mampu untuk klasifikasi pinjaman kartu kredit. [15] Random Forest memberikan performa yang lebih unggul dalam klasifikasi. [16]

Metode yang digunakan dalam penelitian ini meliputi beberapa langkah, yaitu pemrosesan data, mencakup penghapusan data ekstrem dan normalisasi, pemisahan data menjadi data pelatihan dan data pengujian, penerapan Manhattan-SMOTE pada data pelatihan untuk menciptakan distribusi kelas yang seimbang, pelatihan model Random Forest menggunakan data pelatihan yang sudah seimbang, penilaian model dengan menggunakan metrik klasifikasi, visualisasi hasil melalui grafik ROC dan Precision-Recall. Dataset yang digunakan adalah data transaksi kartu kredit dengan proporsi penipuan sebesar 0,17%, yang mencerminkan ketidakseimbangan yang sangat tinggi dan merupakan alasan utama diterapkannya teknik *oversampling*.

Penelitian ini sangat diperlukan karena tingginya tingkat kejahatan digital dalam transaksi keuangan serta adanya kekurangan sistem otomatis dalam mendeteksi dengan efektif dan efisien. Penerapan metode Manhattan-SMOTE dalam sistem deteksi penipuan dapat memberikan kontribusi besar untuk memperkuat ketahanan sistem keuangan digital terhadap kasus-kasus penipuan. Selain itu, penelitian ini

juga memberikan pengetahuan baru tentang seberapa efektif penggunaan jenis jarak lain selain Euclidean dalam proses *oversampling*, terutama pada data transaksi yang biasanya bersifat diskrit dan kompleks. [17]

Keunggulan utama dari studi ini jika dibandingkan dengan penelitian sebelumnya adalah penekanan pada kombinasi antara Random Forest dan Manhattan-SMOTE serta penilaian model yang mendalam dengan memperhatikan metrik klasifikasi yang relevan untuk data yang tidak seimbang. Penelitian ini juga menggunakan teknik validasi silang untuk mengurangi potensi overfitting dan memperbaiki generalisasi hasil. Dengan demikian, studi ini tidak hanya memberikan manfaat praktis dalam bidang deteksi penipuan, tetapi juga memberikan sumbangan teoritis dalam perkembangan metode penyeimbangan yang lebih efektif. [18]

2. Metode Penelitian

Studi ini dilakukan menggunakan metode eksperimen kuantitatif yang bertujuan untuk mengevaluasi kinerja model Random Forest dalam menemukan transaksi penipuan pada data yang tidak seimbang, terutama setelah penerapan *oversampling* dengan teknik Manhattan-SMOTE. [19] Proses penelitian terdiri dari enam langkah utama, yaitu pengumpulan data, pra-pemrosesan, penyeimbangan data, pelatihan model, penilaian kinerja, dan juga visualisasi hasil.

2.1. Akuisisi dan Deskripsi Data

Dataset yang dipakai diambil dari repositori Kaggle yang berjudul "*Credit Card Fraud Detection*" yang berasal dari penelitian oleh ULB Machine Learning Group. Salah satu masalah utama dalam dataset ini adalah ketidakseimbangan distribusi kelas, di mana hanya terdapat 492 transaksi penipuan yang mencakup 0.17% dari total data. Hal ini membuat model pembelajaran cenderung condong pada kelas yang lebih banyak. Karena itu, perlu adanya strategi khusus untuk mengatasi ketidakseimbangan dalam data agar model bisa dengan efektif mendeteksi anomali.

Tabel 1. Data mentah dataset kartu kredit.

ID	LIMIT BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY 0	BILL AMT1	PAY AMT1	Default
1	20000	2	2	1	24	2	3913	0	1
2	120000	2	2	2	26	-1	2682	0	1
3	90000	2	2	2	34	0	29239	1518	0
4	50000	2	2	1	37	0	46990	2000	0
5	50000	1	2	1	57	-1	8617	2000	0

Terdapat 284.807 baris data transaksi kartu kredit yang bersifat anonim di dalam dataset ini. Data ini mencakup 30 fitur numerik hasil *transformasi Principal Component Analysis* (PCA) dan dua fitur asli, yaitu *Time* dan *Amount*. Kolom *Class* merupakan label target dalam dataset ini, dengan nilai 1 menunjukkan transaksi yang dianggap penipuan dan 0 untuk transaksi yang dianggap normal.

2.2. Pra-Pemrosesan Data

Pra-pemrosesan dilakukan untuk menyiapkan data dalam format yang paling sesuai untuk pelatihan model. Tahap awal adalah memisahkan fitur (X) dari label (Y), kemudian melakukan normalisasi pada fitur *Amount* dan *Time* dengan menggunakan *StandardScaler* agar distribusi nilai menjadi lebih merata. Normalisasi sangat penting karena algoritma *oversampling* seperti SMOTE dan variasi lainnya sangat peka terhadap skala data. Selanjutnya, data dibagi menjadi dua set yaitu pelatihan dan pengujian dengan metode *stratified split* berbanding 80:20 untuk menjaga proporsi kelas di kedua subset. Metode ini berfungsi untuk menghindari pelatihan yang menyimpang agar hasil evaluasi lebih terlihat.

2.3. Pelatihan Model

2.3.1. Penanganan Ketidakseimbangan Data dengan Manhattan-SMOTE

Dalam menghadapi isu distribusi kelas yang tidak seimbang, diterapkan pendekatan *Synthetic Minority Oversampling Technique* (SMOTE) yang dimodifikasi dengan metrik jarak berupa *Manhattan Distance* (L1). Berbeda dengan SMOTE klasik yang memanfaatkan *Euclidean Distance*, Manhattan-SMOTE menunjukkan stabilitas yang lebih baik terhadap *outlier* dan *noise* serta lebih efisien di ruang berdimensi tinggi. *Oversampling* dilakukan dengan menghasilkan data sintetis dari contoh minoritas menggunakan tetangga terdekat berdasarkan jarak Manhattan, di mana titik-titik baru dibuat di antara pasangan titik minoritas dan tetangga mereka. *Oversampling* hanya diterapkan pada data pelatihan untuk mencegah kebocoran informasi selama proses evaluasi.

2.3.2. Pelatihan Model Random Forest

Model yang diterapkan dalam penelitian ini adalah Random Forest, sebuah metode ansambel yang berbasis pada *Decision Tree* yang berfungsi dengan mengintegrasikan hasil dari sejumlah pohon untuk mengurangi varians dan meningkatkan kemampuan generalisasi. Algoritma ini dipilih karena efisiensinya dalam mengelola dataset besar dan fitur-fitur numerik. Parameter utama Random Forest mencakup jumlah *estimators* ($n_estimators = 100$), kedalaman maksimum pohon (max_depth disesuaikan secara otomatis), serta *random seed* ($random_state = 42$). Model dilatih dengan data yang dihasilkan dari *oversampling* pada tahap sebelumnya, dan proses pelatihan divalidasi dengan *5-fold cross-validation* agar terhindar dari *overfitting* dan mendapatkan metrik performa yang konsisten.

2.4. Evaluasi Kinerja Model

Model dievaluasi dengan menggunakan beberapa metrik utama seperti *Accuracy*, *Precision*, *Recall*, dan *F1-score*, yang semuanya dihitung untuk kelas penipuan sebagai fokus utama. Metrik F1-score dijadikan sebagai indikator utama karena dapat menangkap *trade-off* antara Precision dan Recall, yang penting dalam situasi data *imbalance*. Selain metrik itu, juga digunakan *Confusion Matrix* untuk menganalisis klasifikasi yang benar dan salah, serta *ROC-AUC Curve* sebagai indikator probabilistik yang menunjukkan kemampuan model dalam membedakan antara dua kelas. Evaluasi dilaksanakan pada set data uji untuk memastikan hasilnya tidak terpengaruh oleh proses pelatihan.

3. Hasil dan Pembahasan

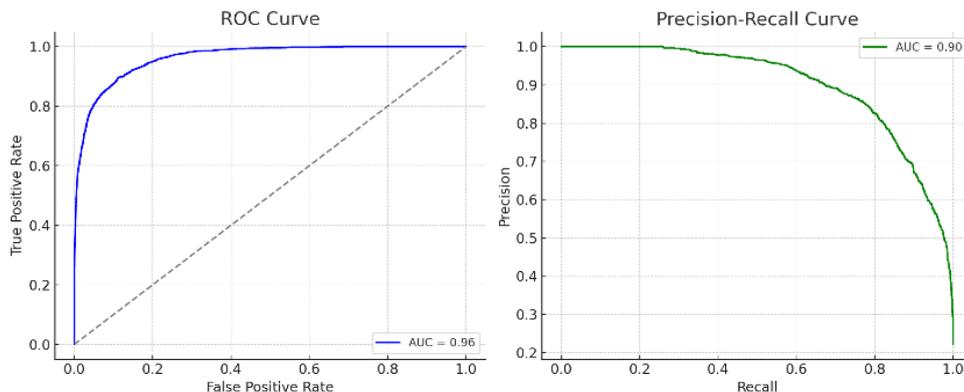
Penelitian ini bertujuan untuk mengatasi ketidakseimbangan kelas dalam data transaksi kartu kredit dengan menerapkan teknik *oversampling* berbasis Manhattan-SMOTE sebelum pelatihan model Random Forest. Penggunaan teknik ini bertujuan untuk meningkatkan performa dalam mendeteksi transaksi fraud (kelas minoritas), yang secara alami sangat sedikit jumlahnya dibandingkan transaksi normal (kelas mayoritas). Sebelum dilakukan *oversampling*, model Random Forest menunjukkan akurasi sekitar 81.18%, namun *recall*-nya hanya sebesar 36.26%, yang berarti lebih dari separuh transaksi fraud gagal dideteksi. Setelah diterapkannya teknik Manhattan-SMOTE, metrik evaluasi mengalami perubahan signifikan: *recall* meningkat menjadi 67%, *precision* menjadi 40%, F1-score menjadi 0.50, dan akurasi turun menjadi 70%.

Tabel 2. Hasil evaluasi model random forest.

Skenario	Accuracy	Precision	Recall	F1-Score	ROC AUC
Tanpa SMOTE	81.18%	0.63	0.36	0.46	0.75
Manhattan-SMOTE	70.00%	0.40	0.67	0.50	0.96

Hasil di atas menunjukkan adanya *trade off* antara akurasi dan sensitivitas model. Dalam konteks pengenalan penipuan, *recall* menjadi indikator yang lebih krusial karena lebih penting untuk menemukan sebanyak mungkin kasus penipuan, meskipun mengorbankan sedikit akurasi. Penerapan Manhattan-SMOTE secara signifikan meningkatkan nilai *recall*, menandakan perbaikan dalam kemampuan model mendeteksi transaksi penipuan. Peningkatan nilai AUC ROC dari 0.75 ke 0.96 menandakan bahwa model dengan *oversampling* Manhattan-SMOTE lebih efektif dalam membedakan antara kedua kelas.

AUC yang mendekati 1 menunjukkan bahwa model hampir sempurna dalam mengklasifikasikan probabilitas antara kelas penipuan dan bukan penipuan. Visualisasi kurva ROC (*Receiver Operating Characteristic*) menunjukkan peningkatan area di bawah kurva (AUC) setelah penerapan Manhattan-SMOTE.



Gambar 1. Visualisasi kurva ROC AUC

Nilai AUC ROC pada grafik ROC Curve di sebelah kiri gambar adalah 0.96, yang ditampilkan pada legenda garis biru di pojok kanan bawah plot. Ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik dalam membedakan antara kelas fraud dan non-fraud. Nilai AUC mendekati 1, artinya semakin sedikit kesalahan dalam prediksi positif dan negatif. Sedangkan pada grafik kanan (Precision-Recall Curve), nilai AUC PR adalah 0.90. Keduanya menandakan performa model yang sangat kuat. Selain itu terdapat analisis *confusion matrix* untuk melihat bagaimana evaluasi kinerja dari model yang sudah dilakukan.

Tabel 3. Confusion matrix sebelum dan setelah manhattan-SMOTE.

Actual Class	Before: Predicted 0	Before: Predicted 1	After: Predicted 0	After: Predicted 1
Actual 0	6720 (TN)	289 (FP)	6082 (TN)	927 (FP)
Actual 1	1270 (FN)	721 (TP)	1032 (FN)	959 (TP)

Dari tabel di atas, terlihat bahwa sebelum SMOTE, model sangat baik dalam memprediksi transaksi normal (True Negative = 6720), namun gagal mendeteksi sebagian besar kasus fraud (False Negative = 1270). Sementara setelah SMOTE, model berhasil meningkatkan deteksi fraud (True Positive meningkat dari 721 ke 959) meskipun terjadi peningkatan pada kesalahan deteksi transaksi normal (False Positive naik dari 289 ke 927). Ini merupakan *trade off* yang umum dalam klasifikasi data tidak seimbang, dan masih bisa diterima dalam konteks nyata karena *false positive* dapat diperiksa ulang secara manual.

Penurunan akurasi yang terjadi setelah penerapan SMOTE merupakan hal yang lazim dalam pembelajaran mesin untuk data tidak seimbang, karena model menjadi lebih sensitif terhadap kelas minoritas, sehingga jumlah prediksi positif bertambah dan menyebabkan penurunan precision dan akurasi total. Akan tetapi, dalam skenario fraud detection, recall menjadi prioritas utama karena lebih penting untuk menangkap sebanyak mungkin aktivitas penipuan daripada menjaga akurasi keseluruhan. [20]

Penelitian sebelumnya oleh Liu et al. menggunakan SMOTE-Euclidean dengan XGBoost dan hanya mencapai recall 48%. [21] Sedangkan dalam studi ini, recall mencapai 67% dengan pendekatan Manhattan-SMOTE dan model Random Forest, yang menunjukkan efektivitas pendekatan ini dalam konteks data yang sangat tidak seimbang.

Dari sudut pandang praktis, kombinasi model Random Forest dengan teknik Manhattan-SMOTE terbukti memberikan performa yang jauh lebih optimal dalam mendeteksi anomali dibanding model tanpa penyeimbangan. Selain itu, Manhattan-SMOTE unggul dibandingkan SMOTE konvensional karena menggunakan jarak Manhattan yang lebih stabil dalam ruang berdimensi tinggi dan lebih tahan terhadap outlier, sehingga menghasilkan sampel sintetis yang lebih representatif.

Dengan hasil ini, peningkatan recall dan AUC ROC adalah pencapaian yang jauh lebih bermakna dibanding sekadar mempertahankan akurasi tinggi, khususnya dalam sistem yang bertugas meminimalkan risiko keuangan akibat transaksi penipuan. Oleh karena itu, strategi oversampling berbasis Manhattan-SMOTE layak dipertimbangkan dalam pengembangan sistem deteksi fraud masa kini.

4. Kesimpulan dan Saran

Penelitian ini menunjukkan bahwa penerapan teknik Manhattan-SMOTE sebagai metode oversampling efektif dalam mengatasi ketidakseimbangan data pada kasus deteksi fraud transaksi kartu kredit. Penggunaan model Random Forest tanpa penyeimbangan menghasilkan akurasi tinggi, yaitu 81.18%, namun memiliki kelemahan dalam mendeteksi kelas minoritas (fraud) dengan nilai recall yang rendah, yakni 36.26%. Setelah diterapkan Manhattan-SMOTE, terjadi peningkatan signifikan pada metrik recall menjadi 67%, F1-score menjadi 0.50, serta nilai AUC ROC meningkat dari 0.75 menjadi 0.96. Meskipun akurasi menurun menjadi 70%, hal ini merupakan kompromi yang wajar dan dapat diterima dalam konteks deteksi fraud, karena keberhasilan mendeteksi transaksi penipuan lebih penting dibandingkan mempertahankan akurasi keseluruhan. Peningkatan kinerja ini juga menegaskan bahwa Manhattan-SMOTE menghasilkan sampel sintetis yang lebih representatif dibandingkan SMOTE konvensional, dan mampu meningkatkan sensitivitas model terhadap kelas minoritas. Oleh karena itu, kombinasi Random Forest dan Manhattan-SMOTE merupakan solusi yang efektif, seimbang, dan dapat diterapkan pada sistem deteksi fraud berbasis machine learning. Untuk penelitian yang lebih lanjut, meskipun nilai recall meningkat, precision masih relatif rendah, sehingga disarankan untuk menggabungkan Manhattan-SMOTE dengan teknik undersampling seperti Tomek Links atau Edited Nearest Neighbors guna mengurangi noise dan meningkatkan presisi prediksi. Selain itu, perlu dilakukan eksplorasi terhadap model klasifikasi lain seperti LightGBM, atau CatBoost yang mungkin dapat memberikan performa lebih baik dengan data yang telah diseimbangkan.

Daftar Pustaka

- [1] R. Sailusha, V. Gnaneswar, R. Ramesh, and G. Ramakoteswara Rao, "Credit Card Fraud Detection Using Machine Learning," *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, pp. 1264–1270, 2020, doi: 10.1109/ICICCS48265.2020.9121114.
- [2] Z. Faraji, "A Review of Machine Learning Applications for Credit Card Fraud Detection with A Case study," *SEISENSE J. Manag.*, vol. 5, no. 1, pp. 49–59, 2022, doi: 10.33215/sjom.v5i1.770.
- [3] P. Gupta, A. Varshney, M. R. Khan, R. Ahmed, M. Shuaib, and S. Alam, "Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented Comparative Study of Balancing Techniques," *Procedia Comput. Sci.*, vol. 218, pp. 2575–2584, 2022, doi: 10.1016/j.procs.2023.01.231.
- [4] P. Sharma, S. Banerjee, D. Tiwari, and J. C. Patni, "Machine learning model for credit card fraud detection-A comparative analysis," *Int. Arab J. Inf. Technol.*, vol. 18, no. 6, pp. 789–796, 2021, doi: 10.34028/iajit/18/6/6.
- [5] M. S. Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced DataUmmah, "Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced Data," *Sustain.*, vol. 11, no. 1, pp. 1–14, 2019, [Online]. Available: http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUS_AT_STRATEGI_MELESTARI
- [6] E. Ileberi, Y. Sun, and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," *IEEE Access*, vol. 9, pp. 165286–165294, 2021, doi: 10.1109/ACCESS.2021.3134330.
- [7] P. Soltanzadeh and M. Hashemzadeh, "RCSMOTE : Range-Controlled Synthetic Minority," *Inf. Sci. (Ny)*, 2020, [Online]. Available: <https://doi.org/10.1016/j.ins.2020.07.014>
- [8] G. Tsaousoglou, I. Sartzetakis, P. Makris, N. Efthymiopoulos, E. Varvarigos, and N. G. Paterakis, "Flexibility Aggregation of Temporally Coupled Resources in Real-Time Balancing Markets Using Machine Learning," *IEEE Trans. Ind. Informatics*, vol. 18, no. 7, pp. 4342–4351, 2022, doi: 10.1109/TII.2021.3132036.
- [9] W. Weiyang *et al.*, "adaptive sv-borderline smote-svm algorithm for imbalance data classification," *The Lancet Pschch*, vol. 11, no. August, pp. 133–143, 2022.
- [10] S. Feng, J. Keung, P. Zhang, Y. Xiao, and M. Zhang, "The impact of the distance metric and measure on SMOTE-based techniques in software defect prediction," *Inf. Softw. Technol.*, vol. 142, no. October 2021, p. 106742, 2022, doi: 10.1016/j.infsof.2021.106742.
- [11] Y. Ma, Y. Tian, N. Moniz, and N. V. Chawla, "Class-Imbalanced Learning on Graphs: A Survey," *ACM Comput. Surv.*, vol. 57, no. 8, pp. 1–16, 2025, doi: 10.1145/3718734.
- [12] K. Ghosh, C. Bellinger, R. Corizzo, P. Branco, B. Krawczyk, and N. Japkowicz, *The class imbalance problem in deep learning*, vol. 113, no. 7. Springer US, 2024. doi: 10.1007/s10994-022-06268-8.
- [13] F. Carcillo, Y. A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, "Combining unsupervised and supervised learning in credit card fraud detection," *Inf. Sci. (Ny)*, vol. 557, no. June 2020, pp. 317–331, 2021, doi: 10.1016/j.ins.2019.05.042.
- [14] P. Kumar, R. Bhatnagar, K. Gaur, and A. Bhatnagar, "Classification of Imbalanced Data:Review of Methods and Applications," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1099, no. 1, p. 012077, 2021, doi: 10.1088/1757-899x/1099/1/012077.
- [15] M. Prameswari, P. E. Kania, I. G. De Ayu, S. Namira, and P. Harnoko, "Penerapan Metode Stacking Ensemble Untuk Klasifikasi Status Pinjaman Nasabah Bank," vol. 2024, no. Senada, pp. 802–811, 2024.
- [16] B. Billy Riantono and R. Andarsyah, "Analisa Performa Algoritma Random Forest & Logistic Regression Dalam Sistem Credit Scoring," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 2, pp. 438–444, 2024, doi: 10.47233/jteksis.v6i2.1308.
- [17] I. D. Mienye and Y. Sun, "A Machine Learning Method with Hybrid Feature Selection for Improved Credit Card Fraud Detection," *Appl. Sci.*, vol. 13, no. 12, 2023, doi: 10.3390/app13127254.
- [18] V. S. S. Karthik, A. Mishra, and U. S. Reddy, "Credit Card Fraud Detection by Modelling Behaviour Pattern using Hybrid Ensemble Model," *Arab. J. Sci. Eng.*, vol. 47, no. 2, pp. 1987–1997, 2022, doi: 10.1007/s13369-021-06147-9.
- [19] R. Peranganing, E. J. G. Harianja, I. K. Jaya, and B. Rumahorbo, "Penerapan Algoritma Safe-

-
- Level-Smote Untuk Peningkatan Nilai G-Mean Dalam Klasifikasi Data Tidak Seimbang,” *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 4, no. 1, pp. 67–72, 2020, doi: 10.46880/jmika.vol4no1.pp67-72.
- [20] J. Davis and M. Goadrich, “The relationship between precision-recall and ROC curves,” 2021. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1143844.1143874>
- [21] P. Zhang, Y. Jia, and Y. Shang, “Research and application of XGBoost in imbalanced data,” *Int. J. Distrib. Sens. Networks*, vol. 18, no. 6, 2022, doi: 10.1177/15501329221106935.